



# 14TH INTERNATIONAL CONFERENCE NOOJ 2020

Book of Abstracts

Virtual Conference  
Zagreb, Croatia  
June 5-7, 2020

Božo Bekavac  
Kristina Kocijan  
Max Silberztein  
Krešimir Šojat (Eds.)

---

## Editors

### **Božo Bekavac**

*Department of Linguistics*

*Faculty of Humanities and Social Sciences*

*University of Zagreb*

*Zagreb, Croatia*

### **Kristina Kocijan**

*Department of Information and Communication Sciences*

*Faculty of Humanities and Social Sciences*

*University of Zagreb*

*Zagreb, Croatia*

### **Max Silberstein**

*Université de Franche-Comté*

*Besançon, France*

### **Krešimir Šojat**

*Department of Linguistics*

*Faculty of Humanities and Social Sciences*

*University of Zagreb*

*Zagreb, Croatia*

---

## Publisher

FF Press

# Organization

---

14<sup>th</sup> Annual NooJ 2020 International Conference was organized as a Virtual conference.

---

## Organizing Institutions



*Department of Information and Communication Sciences  
and Department of Linguistics*  
**Faculty of Humanities and Social Sciences in Zagreb, Croatia**



ELLIADD Laboratory



NooJ Association

---

## Organizing Committee

Kristina Kocijan	University of Zagreb, Croatia
Krešimir Šojat	University of Zagreb, Croatia
Božo Bekavac	University of Zagreb, Croatia
Max Silberztein	Université de Bourgogne Franche-Comté, France

---

## Scientific Committee

Max Silberztein, *Program Committee Chair*, Université de Bourgogne Franche-Comté, France  
Farida Aoughlis Mouloud Mammeri University, Algeria  
Božo Bekavac University of Zagreb, Croatia  
Vincent Bénét Centre d'Etudes Franco Russe CNRS, Russia  
Xavier Blanco Autonomous University of Barcelona, Spain  
Mohamed El Hannache Sidi Mohamed Ben Abdellah University, Fes, Morocco  
Héla Fehri University of Sfax, Tunisia  
Zoe Gavriilidou Democritus Univ. of Thrace, Greece  
Yuras Hetsevlch National Academy of Sciences, Belarus  
Kristina Kocijan University of Zagreb, Croatia  
Philippe Lambert Université de Lorraine, France  
Denis Le Pesant Université Paris 10, France  
Peter Machonis Florida International University, USA

Samir Mbarki IbnTofail University, Morocco  
Slim Mesfar University of Manouba, Tunisia  
Elisabeth Métais Conservatoire National des Arts et Métiers, France  
Mario Monteleone University of Salerno, Italy  
Johanna Monti University of Naples, Italy  
Kamal Naït-Zerrad INALCO, France  
Thierry Poibeau Laboratoire Lattice, CNRS, France  
Jan Radimský University of South Bohemia, Czech Republic  
Andrea Rodrigo University of Rosario, Argentina  
Krešimir Šojat University of Zagreb, Croatia  
François Trouilleux Université Blaise-Pascal, France

## Paraphrasing Emotions in Portuguese

Cristina Mota, Diana Santos and Anabela Barreiro ..... 18

## Text Analysis of Scientific Papers using NooJ: Case Study with Vitamin D Supplementation Debate

Lesia Kaigorodova ..... 20

## Automatic Extraction of French Foods Expression Routine with *NooJ*

Tong Yang ..... 22

## 2. MULTI WORD EXPRESSIONS AND NAMED ENTITIES ..... 25

### Multiword Expressions in the Medical Domain: Who Carries the Domain Specific Meaning

Kristina Kocijan, Krešimir Šojat and Silva Kurolt ..... 26

### Named Entity Recognition with NooJ

Aleksandar Petrovski ..... 28

### Optimization of Portuguese Named Entity Recognition and Classification by Combining Local Grammars and Conditional Random Fields Trained with Parsed Corpora

Diego Válio Antunes Alves, Božo Bekavac and Marko Tadić ..... 30

### Automatic Recognition and Translation of Tunisian Dialect Named Entities into Modern Standard Arabic

Roua Torjmen and Kais Haddar ..... 32

## 3. TEXT MINING AND QUESTION ANSWERING ..... 35

### Annotation of Cause - Result Questions in Standard Arabic Using Syntactic Grammars

Bessaies Essia, Mesfar Slim and Ben Ghazela Henda ..... 36

### A Bottom-Up Approach for Moroccan Legal Ontology Learning from Arabic Texts

Kautar Belhoucine, Mohammed Mouchid, Aziz Mouloudi and Samir Mbarki ..... 38

### Answering Arabic Complex Question

Sondes Dardour, Hela Fehri and Kais Haddar ..... 40

### Development of a Question-Answering System in the Legal Field Based on Ontologies

Ismahane Kourtin, Aziz Mouloudi and Samir Mbarki ..... 42

## 4. AI, LEXICONS AND DICTIONARIES ..... 45

### NooJ for Artificial Intelligence: an Anthropic Approach

Mario Monteleone ..... 46

# Text Analysis of Scientific Papers using NooJ: Case Study with Vitamin D Supplementation Debate

Lesia  
Kaigorodova

## Abstract

This work is dedicated to text mining of scientific papers with NooJ regarding vitamin D consumption, probably the most popular topic nowadays in the era of supplements.

We will try to design a system that would provide answers to such questions as:

- do we benefit from vitamin D intake or is it harmful, and what are the restrictions?
- how does supplementation work in the presence of a particular disease?
- can we provide medical practitioners with up-to-date knowledge in the domain?

The research can be extended to address other queries that may refer to medicine as well.

The data for the research was collected from the PubMed database of the NCBI (United States National Center for Biotechnology Information) website. It has the greatest selection of articles in the field of medical domain. Overall 83 077 articles were scraped from the PubMed database that refer to the keyword “vitamin D”. The data is comprised of a title, list of authors, date, journal, citation and an abstract. Some abstracts contain more structured information in the form of sections such as “conclusion”, “objectives”, “method”, “results”, “design”, “setting”.

Automatic Opinion Analysis is used in order to address the debate that has occurred in the modern world regarding vitamin D consumption. It



UIIP, National Academy  
of Sciences of Belarus  
[Minsk, Belarus]



lesia.piatrouskaya  
@gmail.com



*Automatic opinion  
analysis  
NLP  
Text mining  
Vitamin D  
Medical domain  
Belarus language  
NooJ*

## References

Hetseвич, Yu. **Grammars for Sentence into Phrase Segmentation: Punctuation Level**. In *Automatic Processing of Natural-Language Electronic Texts with NooJ: 9th International Conference, NooJ 2015, Minsk, Belarus, June 11-13, 2015, Revised Selected Papers* / ed. T. Okrut, Y. Hetseвич, M. Silberstein, H. Stanislavenka. — Springer

International Publishing, 2016.  
pp. 74-82.

Silberztein Max, 2018. **NooJ Manual**. Available for download at: <http://www.nooj-association.org>.

should be mentioned that specific terminology and language patterns should be taken into consideration in order to process scientific papers.

The concept for "vitamin D" is introduced as well as positive and negative features for the outcome of vitamin D supplementation. For example, the "vitamin D" concept comprises {vitamin D, cholecalciferol, D2, D3, 1,25-dihydroxyvitamin, 25(OH), VD, ...}. For "positive features" of the outcome it may be {promising, good, help, safe, improve, stabilize, regulate, positive, progress, ...}. For the "negative features" it may be {negative, dangerous, unsafe, side effect, cautious, abnormal, harmful, trigger, worse, ...}. It should be noted that positive and negative features should be located and processed in the text according to its context since it can be negated and hence represent the opposite.

The grammars are constructed to make all the introduced concepts work together in order to extract meaning from the abstracts, i.e. positive, negative or neutral outcome of the vitamin D intake. Other additional information such as sections "objectives", "conclusion", etc. is analyzed for the inference of particular knowledge.

With NooJ it may be possible to design a knowledge-based system for the available information in the domain of medicine. It could be helpful for both medical practitioners to make information-driven decisions and for individuals who are desperate to find the truth in the abundance of contradicting points of view.